Multi-objective local search for mining Pittsburgh classification rules

Julie Jacques, Julien Taillard, David Delerue, Laetitia Jourdan, Clarisse Dhaenens

Multi-objective local search for mining Pittsburgh classification rules

J. Jacques[1,3], J. Taillard[1], D. Delerue[1], L. Jourdan[2,3], and C. Dhaenens[2,3]

[1] Société ALICANTE, 50 Rue Philippe de Girard, 59113 Seclin, France
julie.jacques@alicante.fr, julien.taillard@alicante.fr, david.delerue@alicante.fr
[2] INRIA Lille Nord Europe, 40 Av. Halley, 59650 Villeneuve d'Ascq, France
[3] LIFL, Université de Lille 1, Bât. M3, 59655 Villeneuve d'Ascq cedex, France
laetitia.jourdan@inria.fr,clarisse.dhaenens@lifl.fr

1 Context

Classification rule mining aims to extract, from known data, rules predicting a chosen fact – called *class*, which can be for example *cardiovascular risk* in a medical context. The following rule is an example of what can be extracted:

$$\text{diabetes and age} > 70 \rightarrow \text{cardiovascular risk}$$

Training data is composed of observations (sells, credit approvals, patients...), each of them having some attributes (e.g.: items sold, salary, diagnosis...). The obtained prediction rules are understandable, in contrast to models found by *neural networks* or *support vector machines*. *Decision trees* (C4.5, Oblique-DT, DT-GA) are another popular method but have a drawback when multiple contexts can explain a same class, since they are not able to find overlapping rules. This can happen in medical data, e.g. *cardiovascular risk* is higher when both *high blood pressure* and *diabetes* are found than when only *diabetes* or *high blood pressure* is found.

2 Multi-objective local search for classification rule mining

A large number of available rules can be generated from a small number of attributes; exploring all of them to extract only the interesting ones is a costly task, even impossible when a lot of attributes are available. Therefore combinatorial optimization methods are candidate methods to deal with rule mining: they are able to deal with large search spaces. Two major rule encodings are available; *Michigan* is the common one, where each solution is a single rule. In *Pittsburgh* encoding each solution is a set of rules. This encoding increases the search space and adds complexity for mutation and cross-over operators but allows finding complementary rules, like in medical data. *Accuracy* is often used as a fitness function. *Accuracy* measure counts good classifications provided by a rule: *true positives* and *false negatives*, over all classifications given by this rule. A rule with an *accuracy* of 1 makes no wrong classification, while a rule with an *accuracy* of 0.7 is wrong for 30% of observations. Another popular criterion is *Minimum description length* (MDL) principle [4], an application of Occam's razor: given two equivalent rules, the simplest rule (the shortest) must be preferred.

Multi-objective approach can handle mining rules on multiple criteria, obtaining rules having both good performance and simplicity. Many multi-objective methods were proposed for rule mining and most of them are detailed in Srinivasan and Ramkrishnan's review [5]. Methods using an aggregation of objectives can give interesting results, like *learning classifier systems (LCS)*, including GAssist [1] and XCS [6]. Most of them are based on *genetic algorithms*, especially NSGA-II. But one of them uses GRASP which is a greedy algorithm.

We propose a multi-objective model, based on 2 criteria: maximizing *accuracy* and minimizing the number of terms (MDL principle). We implemented a *Dominance-based multi-objective local search* (DMLS), which is a population-based local search algorithm dedicated to multi-objective and has proven to give at least as good results as NSGA-II on several problems [3]. Moreover, DMLS is easier to parameter than a GA and does not need any cross-over operator. Our solution encoding is Pittsburgh. Each of our rules-sets contains only partial classification rules (e.g.: only rules predicting *cardiovascular risk*), avoiding rule inconsistency.

3 Results

3.1 Experimentations

Fernández et al compared 22 state-of-the-art classification rule mining algorithms on 30 datasets, providing their obtained *accuracy* on each dataset [2]. We compared our method to the results they obtained. In the proposed datasets, we selected 5 of them with less continuous attributes since our model was designed to handle discrete attributes. When continuous attributes were available (*crx*, *hea* and *hep* datasets), we discretized each attribute in 10 bins. According to Fernández et al protocol, our algorithm was run 25 times for each dataset. Datasets are split into 5-fold cross-validation: 20% of observations in each fold. Then 4 folds are used for training, 1 for evaluation. For each available partition, the algorithm was run 5 times. After each run, obtained rule sets are merged into one rule set, on which we can compute *accuracy*.
Results are available in Table 3.1. We selected among the 22 available, the 10 algorithms giving the best results. For each dataset (*bre*, *crx*, *hea*, *hep*, *tic*) we selected the best obtained *accuracy* by these algorithms. Then we computed the relative error to the best for each algorithm. A value of 0 indicates the algorithm that obtained the best *accuracy*. Our results are available in the first column.

	MOCA	**XCS**	**SIA**	**Oblique-DT**	**CORE**	**Gassist**	**OCEC**	**DT-GA**	**HIDER**	**C4.5**
bre	**0.0000**	0.0074	0.0795	0.1672	0.0325	0.0200	0.1333	0.0278	**0.0000**	0.0144
crx	0.0008	0.0067	0.2404	0.0727	0.0411	0.0060	**0.0000**	0.0043	0.0625	0.0124
hea	0.0686	0.0194	0.1563	0.0759	0.0675	**0.0000**	0.0259	0.0352	0.0962	0.0056
hep	0.0500	**0.0000**	0.0667	0.0750	0.1111	0.0472	0.1222	0.0944	0.0639	0.1250
tic	0.2650	0.1448	**0.0000**	0.0971	0.2965	0.0467	0.1860	0.1716	0.2991	0.1400

Table 1. Relative error to the best for some state-of-the-art classification rule mining algorithms

MOCA: Multi-Objective Classifier Algorithm, SIA: Supervised Inductive Algorithm, Oblique-DT: Oblique Decision Tree, CORE: CO-Evolutionary Rule Extractor, GAssist: Genetic Algorithms based claSSIfier sySTem, OCEC: Organizational Co-Evolutionary algorithm for Classification, DT-GA: Hybrid Decision Tree - Genetic Algorithm, HIDER: HIerarchical DEcision Rules

3.2 Discussion

Table 3.1 shows that our method obtained the best solutions on *bre* dataset. When outperformed, it is each time by different algorithms, moreover on *crx* the obtained *accuracy* is very close to the best. Furthermore, these results are obtained with a local search, which is easier to configure than most of evaluated algorithms that are GA. Bad results on *tic* and *hea* datasets suggest neighborhood operators may be weak regarding those used by GA: GAssist with a similar model but more operators obtained better results. A future work could focus on improving our operators.

References

1. Bacardit, J.; Burke, E. K. and Krasnogor, N.: Improving the scalability of rule-based evolutionary learning. Memetic Computing, Springer Berlin / Heidelberg, 2008, 1, 55-67
2. Fernández, A.; García, S.; Luengo, J.; Bernadó-Mansilla, E. and Herrera, F.: Genetics-Based Machine Learning for Rule Induction: State of the Art, Taxonomy, and Comparative Study. IEEE Transactions on Evolutionary Computation, 2010, 14, 913-941
3. Liefooghe, A.; Humeau, J.; Mesmoudi, S.; Jourdan, L. and Talbi, E.-G.: On dominance-based multi-objective local search: design, implementation and experimental analysis on scheduling and traveling salesman problems. Journal of Heuristics, Springer Netherlands, 2012, 18, 317-352
4. Rissanen, J.: Modeling by shortest data description. Automatica, 1978, 14, 465 - 471
5. Srinivasan, S. and Ramakrishnan, S.: Evolutionary multi objective optimization for rule mining: a review. Artificial Intelligence Review, Springer Netherlands, 2011, 36, 205-248
6. Wilson, S. W.: Classifier fitness based on accuracy. Evol. Comput., MIT Press, 1995, 3, 149-175